Promoting Cultural Responsivity and Student Engagement through Double Check Coaching of

Classroom Teachers: An Efficacy Study

Catherine P. Bradshaw

University of Virginia

Elise T. Pas

Johns Hopkins University

Jessika Bottiani, Katrina J. Debnam,

University of Virginia

Wendy Reinke, Keith Herman,

University of Missouri

Michael Rosenberg,

SUNY New Paltz

School Psychology Review

Authors' Notes

Correspondence concerning this article should be addressed to Catherine Bradshaw, Curry School of Education, University of Virginia. Charlottesville, VA, 22904-4260. E-mail: CPB8g@virginia.edu

Abstract

This paper presents findings from a randomized controlled trial (RCT) testing the impact of a novel coaching approach utilized as one element of the Double Check cultural responsivity and student engagement model. The RCT included 158 elementary and middle school teachers randomized to receive coaching or serve as comparisons; all participating teachers were exposed to school-wide professional development activities. Pre-post non-experimental comparisons indicated improvements in self-reported culturally responsive behavior management and self-efficacy for teachers in both conditions following professional development exposure. With regard to the experimental findings, trained observers recorded significantly more teacher proactive behavior management and anticipation of student problems, higher student cooperation, less student non-cooperation, and less disruptive behaviors in classrooms led by coached teachers relative to comparison teachers. Taken together, the findings suggest the potential promise of coaching combined with school-wide professional development for improving classroom management practices and possibly reducing office discipline referrals among Black students.

Keywords: intervention, behavior, prevention, school discipline, school engagement, student-teacher relationships

Promoting Cultural Responsivity and Student Engagement through Double Check

Coaching of Classroom Teachers: An Efficacy Study

Office discipline referrals, suspensions, and expulsions are considered exclusionary disciplinary practices because they remove and exclude students from schools and classroom learning time as a punitive consequence to an infraction of school or classroom rules or expectations. Studies consistently show that Black youth are more likely to experience exclusionary discipline in elementary and secondary school than their peers. Although Latino and Native American secondary school students also experience disproportionately high rates of exclusionary discipline relative to their White peers (Losen, Hodson, Keith, Morrison, & Belway, 2015; Losen, Hodson, Ee, & Martinez, 2015), the discipline gap for Black relative to White students is by far the largest (Bradshaw, Mitchell, O'Brennan, & Leaf, 2010; Losen, Hodson, Ee, et al., 2015; Losen, Hodson, Keith et al., 2015; Porowski, O'Conner, & Passa, 2014; Skiba, Horner, Chung, Rausch, May, & Tobin, 2011). Furthermore, Black youth experience harsher punishments (e.g., higher average number of days of suspension) for similar violations of school rules than their White peers (see Bradshaw, Mitchell, O'Brennan et al., 2010; Losen, Hodson, Ee et al., 2015; Skiba et al., 2011; Skiba, Chung, Trachok, Baker, Sheya, & Hughes, 2014; Skiba, Michael, Nardo, & Peterson, 2002; Skiba, Shure, & Williams, 2012).

Unfortunately, this problem appears to be worsening over time, as data from the U.S. Department of Education's Office of Civil Rights suggest the risk of out-of-school suspension has increased disproportionately for Black youth (Losen, Hodson, Keith, et al., 2015; U.S. Department of Education, Office of Civil Rights, 2014, March). Moreover, there are numerous detrimental effects of removal and exclusion from school and classroom environments, including increased risk of subsequent school dropout and contact with the juvenile justice system, even

after controlling for other socio-demographic characteristics (Bradshaw, O'Brennan, &

McNeely, 2008; Debnam, Bottiani, & Bradshaw, 2017; Fabelo et al., 2011). These findings

highlight the contribution of exclusionary school discipline practices to the school-to-prison

pipeline (Skiba et al., 2011).

Despite mounting concern about the issue of disproportionality (Skiba et al., 2012), there

has been limited empirical examination of programs or interventions that actually reduce these

gaps. For example, a recent systematic literature review on in-service professional development

models of culturally responsive practices for educators revealed that only 10 out of the 179 peer-

reviewed articles reported empirical results from an intervention study (Bottiani, Larson,

Debnam, Bischoff, & Bradshaw, 2017). Rather, the majority of the 179 studies reviewed focused

on teacher characteristics or theory-based recommended strategies, rather than outcomes-focused

intervention studies. Furthermore, none of the studies met necessary standards of evidence, as

defined by the Department of Education's What Works Clearinghouse, to substantiate

conclusions regarding promising and evidence-based strategies, because they did not employ

rigorous design features allowing for causal inferences to be drawn (see Bottiani et al., 2017).

Taken together, these findings highlight the need for more rigorous research on the outcomes

associated with in-service professional development related to cultural responsivity.

The current study aimed to address this gap by developing and testing, in part with a

randomized trial, a newly-developed preventive intervention called Double Check (Bradshaw &

Rosenberg, 2018). Double Check is a professional development and coaching framework that

builds on School-Wide Positive Behavioral Interventions and Supports (SW-PBIS) to help

teachers enhance five core components of culturally responsive practices. The overarching goal

of Double Check is to address the overrepresentation of students of color in disciplinary

referrals, suspensions, and special education referrals. The current study reported pre-post non-experimental findings for the school-wide professional development, as well as results from a teacher-level randomized controlled trial experimentally testing the unique effects of the coaching element of the broader Double Check model on teacher and student outcomes. In the sections below, we provide an overview of the five core domains of culturally responsive practices which serve as the conceptual foundation for the Double Check model and the corresponding professional development and coaching activities. Additional information on the theory of change process underlying the full Double Check intervention is provided in the logic model (see Figure 1).

Overview of the Double Check Model

The Double Check preventive intervention included three core components: (a) augmentations to the tier 1 elements of SW-PBIS (Sugai & Horner, 2002, 2006); (b) a set of five 60-minute professional development trainings addressing five core domains of culturally responsive practice, which we refer to as the Double Check CARES domains (i.e., **C**onnection to the curriculum, **A**uthentic relationships, **R**eflective thinking, **E**ffective communication, and **S**ensitivity to students' culture; Bradshaw & Rosenberg, 2018; Bottiani, Bradshaw, Rosenberg, Hershfeldt, Pell, & Debnam, 2012; Hershfeldt, Sechrest, Pell, Rosenberg, Bradshaw, & Leaf, 2009); and (c) individual classroom coaching using an adapted version of the Classroom Check-Up (CCU; Reinke, 2006; Reinke, Herman, & Sprick, 2011; Reinke, Lewis-Palmer, & Merrell, 2008), which utilized structured problem-solving to facilitate changes in teacher practices (see Pas, Larson, Reinke, Herman, & Bradshaw, 2016). Together, these activities were intended to increase staff cultural responsivity and classroom management skills and, ultimately reduce disproportionality in teachers' use of exclusionary discipline (see Figure 1). In turn, student

engagement, student behavior, and academic performance are expected to improve. Although all three elements of the Double Check model are critical, by design, the largest proportion of coaches' effort is focused on coaching of individual teachers, followed by the professional development, and then by the SW-PBIS supports. As such, the SW-PBIS elements serve as a foundation for, but not main focus of, the Double Check model. Each of the three components of Double Check is described in detail below, with relevant research supporting its inclusion in the full Double Check model.

SW-PBIS Tier 1 Augmentations. A core assumption of the Double Check model is that clear and consistent school-wide and classroom-based management practices are a necessary, but not sufficient condition for culturally responsive instruction, equity, and student engagement. As a result, the model builds on foundational tier 1 elements of SW-PBIS. Consistent with the model outlined by Sugai and Horner (2002, 2006), SW-PBIS is a comprehensive, data-driven process which emphasizes consistent implementation of evidence-based programs and school-wide practices to improve school climate and reduce discipline problems. SW-PBIS is premised upon the public health framework for preventive intervention and mental health promotion (O'Connell, Boat, & Warner, 2009) through which the universal or school-wide component is delivered at tier 1, allowing for evidence-based selective/targeted and indicated/intensive preventive interventions to be delivered through a multi-tiered system of supports at tiers 2 and 3. Although SW-PBIS has demonstrated success in improving school climate and reducing office disciplinary referral rates overall (Bradshaw, Koth, Bevans, Ialongo, & Leaf, 2008; Bradshaw, Mitchell, & Leaf, 2010; Bradshaw, Waasdorp, & Leaf, 2012; Horner et al., 2009), several studies have shown that racial disparities in school discipline practices affecting Black students remained relatively unchanged in schools implementing SW-PBIS (Bradshaw, Mitchell,

O'Brennan et al., 2010; Bradshaw, Waasdorp, & Leaf, 2015; Vincent, Randall, Cartledge, Tobin, & Swain-Bradway, 2011). As a result, these foundational school-wide PBIS activities were augmented for the purposes of Double Check; this included additional coaching provided to the SW-PBIS team regarding data-based decision-making related to the review, disaggregation, and use of ODR data as well as support for other school-wide approaches to increase student engagement and the use of culturally responsive disciplinary approaches.

Double Check Professional Development Model. The Double Check professional development series was created to map onto the five CARES domains described below (for additional information see Bottiani et al., 2012; Bradshaw & Rosenberg, 2018; Hershfeldt et al., 2009). Consistent with the behavioral theoretical focus of the Double Check model (Hershfeldt et al., 2009) and PBIS (Sugai & Horner, 2002, 2006), there is a specific emphasis within each professional development on skill development through the provision of concrete strategies and practices that teachers can utilize within their classrooms, rather than a sole focus on awareness-raising. Specifically, *Connection to the Curriculum* refers to tangibly linking all elements of lesson planning and classroom instruction to students' cultural backgrounds, and promoting a partnership between teacher and student in the goal of mastering the material (Gay, 2002; Ross, Kamman, & Coady, 2008). Examples include visible images reflecting cultural values and the selection of learning activities that reflect and acknowledge students' backgrounds, cultural norms, families, and communities. *Authentic Relationships* between educators and students are characterized by tangible evidence of warmth, caring, and trust. There is a focus in this professional development on strategies for revealing the teachers' enthusiasm, caring, and willingness to help students meet academic and behavioral challenges, as well as ways to learn more deeply about their students. Positive relationships between students and teachers help

students adjust to the pressures of the classroom, promote social competence, and are associated with fewer behavior problems (Debnam, Lindstrom Johnson, Waasdorp, & Bradshaw, 2014; Hurd, Hussain, & Bradshaw, 2015). *Reflective Thinking* relates to the examination of one's own social, cultural, and class membership, and how these factors interact with the group memberships of students as well as teachers' responses to specific student behaviors in the classroom. Strategies of focus in this professional development include reflective thinking about situations that arise within the classroom, considering the functions of student behaviors, and mindfulness for the purpose of stress management. Reflective practices such as mindfulness and perspective-taking have the potential to slow down and interrupt the often automatic processes or unconscious biases and responses teachers may display in responding to student behaviors (Bottiani et al., 2012; Jennings, 2015; Jennings, Frank, Snowberg, Coccia, & Greenberg, 2013; Wang, Tai, Ku, & Galinsky, 2014). *Effective Communication* refers to recognizing the distinct interactive styles of students from different backgrounds and knowing how to respond differentially. Consistent with the behavioral focus of the Double Check model, this professional development focuses on: (a) understanding the communicative function of the student's different behaviors; (b) encouraging consistent interactions that reflect credibility, civility, and respect; (c) limiting judgmental verbal interactions directed toward the student; and (d) facility with code-switching to mitigate the burden associated with frequent adjustment to varying standards of communication and behavior in different contexts (e.g., home, neighborhood, school; Day-Vines & Day-Hairston, 2005; Gay, 2010; Richards, Brown, & Forde, 2007). Finally, *Sensitivity to Students' Culture* requires both an understanding that students may present behaviors that do not align with the school's behavioral expectations, and that these behaviors reflect differences rather than deficits (Day-Vines & Day-Hairston, 2005). The professional development focused in

this area aims to help teachers build a classroom community which highlights the strengths of each student's cultural background.

Classroom Check-Up (CCU; Reinke, 2006; Reinke et al., 2011; Reinke et al., 2008). The original CCU coaching model is a data-based decision making process whereby the coach and teacher go through five clear steps; several of the original CCU materials (e.g., interview, data collection) were augmented to reflect Double Check's focus on cultural responsivity. These adapted steps include: (a) initial rapport building through an interview focused on teacher strengths, classroom management and culturally responsive practices the teacher currently engages in and would like to change or enhance, and previous experiences with coaching; (b) data collection via a teacher-reported survey called the Classroom Ecology Checklist, which assesses teachers' perceptions of their use of targeted classroom management and culturally responsive strategies, as well as three coach-conducted classroom observations to quantitatively assess overall classroom climate and management and qualitatively assess each of the CARES elements; (c) feedback to the teacher regarding areas of relative strength and weakness (i.e., synthesized from the interview, checklist, and observations); (d) collaborative goal setting of the adoption or increase of a specific strategy related to positive behavior supports and CARES and a plan for how that would be accomplished; and (e) implementation with progress monitoring and follow-up feedback from the coach, who fades out support after two follow-up meetings. The coach uses a communication technique called motivational interviewing (Miller & Rollnick, 2002) throughout this process as a means to empower teachers to adopt and sustain their use of new or improved classroom behavior management and culturally responsive practices. Motivational interviewing was initially incorporated into this coaching approach as a way to address teachers' ambivalence about changing their practice (Miller & Rollnick, 2002; Reinke et

al., 2011), but given the concern with implementation fidelity in schools and when addressing a more sensitive topic like culturally responsive practices, it is particularly important. For a more complete description of the original CCU (i.e., focused only on classroom management), see Reinke et al. (2011) and for additional detail on the coaching materials that align with the five Double Check CARES domains see Bradshaw and Rosenberg (2018) and Pas et al. (2016).

Overview of the Current Study

The primary purpose of the current study was to conduct a preliminary efficacy test of the Double Check coaching model, using a within-school, teacher-level randomized controlled trial (RCT) design in 12 schools (i.e., 6 middle and 6 elementary). In a separate analysis, we examined non-experimental pre-post changes for the full sample of teachers following the professional development activities. All 12 schools had been previously trained in the SW-PBIS (Sugai & Horner, 2002, 2006) model and were implementing the tier 1 SW-PBIS supports with high fidelity according to the School-wide Evaluation Tool (SET; Horner et al., 2004). In addition, all school staff (including teaching and non-teaching staff) participated in the five-part Double Check professional development series; therefore, only a non-experimental pre-post study was feasible for estimating the effects of the professional development. However, the teacher-level RCT design enabled us to isolate the impact of the Double Check coaching model over and above the effect of the school-wide professional development and SW-PBIS augmentations. Specifically, we were particularly interested in determining the unique effects of the coaching, beyond the other elements of the Double Check model, given the additional cost and burden associated with coaching of individual teachers, both in terms of teacher time and coach time. In fact, to our knowledge, there has been no rigorous test of coaching as a professional development for promoting equity or culturally responsive instruction (Bottiani et

al., 2017). As such, the primary aim of this study was to determine whether the randomly

assigned coaching resulted in a "value added", over and above the pre-post (non-experimental)

changes we anticipated would occur school wide following the implementation of SW-PBIS and

the delivery of the Double Check professional development. Although our primary hypotheses

focused on improvements for teachers randomized to receive coaching, thereby leveraging data

from the teacher-level RCT design, we were also interested in the pre-post non-experimental

comparisons for the full sample of participating teachers; this latter set of pre-post comparisons

enabled us to contextualize other changes that may have occurred over the course of the one-year

project in response to the school-wide and professional development efforts.

Given this was the first test of this novel intervention, we examined impacts on a range of

possible outcomes and leveraged data from a variety of sources. Specifically, our primary

outcomes of interest were the behavioral indictors, including ODR data submitted formally

through a school-wide data system (i.e., School-wide Information System [SWIS]; Irvin, Tobin,

Sprague, Sugai, & Vincent, 2004; May et al., 2003) and observations of classrooms conducted

by external trained assessors who were unaware of the teachers' randomization status. A

secondary set of outcomes included teacher self-reported behaviors and attitudes (e.g., stress,

self-efficacy), which also enabled us to be sensitive to potential iatrogenic effects (e.g., increased

stress, reduced self-efficacy) which could occur when directly discussing sensitive topics such as

equity, culture, and disproportionality (Tatum, 2017).

Method

A total of 12 schools (6 middle and 6 elementary) from a single Maryland public school

system participated in the project (see Table 1 for school demographics). The schools had an

average enrollment of 594.33 students who were racially and ethnically diverse (i.e., 37.93% of

students were Black, 36.17% were White, and 14.49% were Hispanic). Approximately 56.83%

of students received free and reduced priced meals. Across the 12 schools, 158 eligible

classroom teachers (i.e., those teaching core instructional courses) consented to randomization

within the RCT. The majority of the randomized teacher sample taught in middle school (i.e.,

65.8%), were White (75.3%) and were female (i.e., 85.4%). Among those in the randomized

sample, 100 were randomly assigned to receive coaching; again, the majority of whom (i.e.,

66%) taught in a middle school (i.e., 34% taught in an elementary school) and were female (i.e.,

88% of teachers). T-tests indicated no significant differences between the demographics of

teachers randomly assigned to the coached condition vs. the comparison (i.e., non-coached)

condition, thereby demonstrating balance across the two conditions. Participant demographic

characteristics and baseline (i.e., Time 1) data are reported in Table 2.

Table 1. *School demographic data at Time 1*

School Demographics	*M*	*SD*	Min	Max
Student Enrollment	594.33	214.21	223	1049
% White	36.17	17.01	3.2	60.2
% Black	37.93	17.67	19.2	80.7
% Hispanic	14.49	5.87	6.9	25.9
% English Language Learners	7.82	2.52	5.0	12.6
% Receiving FARMS	56.83	19.16	32.1	87.8
Math Proficiency Rate	73.93	11.05	54.1	90.5
Reading Proficiency Rate	83.80	4.60	73.3	90.2
Suspension Rate	13.97	9.22	0.5	27.2
Years since PBIS Training	7.67	3.63	1	12

Note. Some percentages do not add to 100% because of missing data. FARMS = percent of students receiving free and reduced priced meals.

Table 2. *Demographics and Time 1 means for the full sample of teachers randomized to either the coached or comparison condition*

	Full Sample (N = 158)		Coached (N = 100)		Comparison (N = 58)	
Teacher Demographics (N = 158)	N	%	N	%	N	%
Female	135	85.4	86	86.0	47	81.0
White Race	119	75.3	74	74.0	45	77.6
Black Race	19	12.0	14	14.0	5	8.6
Other Race	14	8.9	9	9.0	5	8.6
Early Career (20-30)	53	33.5	33	33.0	20	34.5
Middle School Teacher	104	65.8	66	66.0	38	65.5
Time 1 Means	M	SD	M	SD	M	SD
Office Disciplinary Referrals						
Total ODRs	3.5	5.4	3.4	5.8	3.6	5.0
ODRs - Black Students	2.2	3.9	2.2	4.2	2.3	3.4
ODRs - White Students	0.7	1.5	0.7	1.5	0.8	1.6
ASSIST Tallies						
Proactive Behavior Expectations	7.0	5.3	7.2	5.3	6.7	5.2
Approvals	6.4	7.0	6.2	6.9	6.6	7.2
Disapprovals	0.4	1.1	0.3	0.9	0.4	1.4
Opportunities to Respond	21.6	15.0	22.5	15.1	20.0	15.0
Non-cooperation	0.3	0.9	0.2	0.6	0.6	1.2
Disruptive Behaviors	12.1	8.9	11.7	8.4	12.9	9.7
ASSIST Global Ratings						
Cultural Responsiveness	1.4	1.1	1.4	1.0	1.4	1.1
Teacher Direction & Influence	3.6	0.5	3.6	0.5	3.6	0.6
Teacher Responsiveness	3.3	0.8	3.3	0.8	3.3	0.8
Teacher Proactive Beh. Mgmt.	3.2	0.6	3.2	0.5	3.2	0.7
Meaningful Participation	2.9	1.0	2.9	1.0	2.8	1.0
Student Cooperation	3.3	0.7	3.4	0.7	3.3	0.8
Student Social Disruptions	1.0	0.5	0.9	0.5	1.1	0.5
Teacher Self-Report						
CRT Efficacy	4.3	0.5	4.2	0.5	4.4	0.5
Multicultural Efficacy	4.6	0.5	4.6	0.4	4.7	0.5
Behavioral Management Efficacy	4.5	0.7	4.4	0.7	4.5	0.6
Stress	3.7	1.0	3.8	1.0	3.6	1.1

Note. Some percentages do not add to 100% because of missing data.

Specifically, teachers were matched within schools and randomly assigned to receive the Double Check coaching or to serve as comparisons within the context of the whole-school Double Check program, whereby all school staff members (i.e., regardless of teaching role and randomization status) were exposed to the five-part, school-wide Double Check professional development trainings; this was in addition to the schools' universal implementation of SW-PBIS augmented by additional support from the Double Check coach. We randomized a slightly higher proportion of teachers to the coached condition in order to conduct a descriptive sub-study exploring factors associated with quality coaching, the findings of which are summarized in Pas et al. (2016).

Procedure

Schools' participation in the Double Check project was voluntary. The participating school district hosted principal meetings to obtain interest and school-level commitment to the project. Principals were asked to sign commitment letters indicating their willingness to participate in the study, including having teachers randomized to receive Double Check classroom coaching. Written consent to participate in the coaching was solicited from all eligible classroom teachers; the consenting teachers were then randomized to receive Double Check classroom coaching or serve as comparison teachers. Approximately 42.5% of the eligible teachers consented to participate in the randomized study.

All consenting teachers were asked to complete self-report surveys at the start (i.e., pre-test/baseline/fall) of the study prior to randomization and at the end (i.e., post-test/spring) of the school year, and for observations to be conducted in their classrooms at these two times. Self-report surveys were completed via a secure online survey system. All consenting teachers received a $10 gift card at pre-test and again at post-test in appreciation of their time spent

completing the surveys. The teacher participation rate in the data collection among consenting

teachers at pre-test was 96% for the teacher self-report survey and 99% for the classroom

observations; at post-test, participation rates were 95% for the teacher self-report survey and

89% for the classroom observations. The researchers' Institutional Review Board approved this

study.

School-wide Professional Development. All teachers in the 12 schools were invited by

project staff to attend the five Double Check professional development sessions and school

leadership helped to facilitate attendance. With regard to fidelity (i.e., dosage), among the 158

consented teachers, participating teachers, on average, attended 73.10% of the professional

development sessions, or 3.65 (out of 5) sessions. About two-thirds (62.6%) of the teachers

attended 4 or 5 of five professional development sessions. T-tests indicated no significant

differences on session-specific or cumulative attendance rates between intervention and

comparison teachers.

Coaching. Coaching was provided only to the consenting teachers who were randomized

to the intervention condition, relative to a non-coached comparison condition. Specifically, the

goal of the Double Check CCU coaching was to provide on-going support for teachers' adoption

of classroom management and culturally responsive practices taught in the professional

development, by applying a structured problem-solving approach (for additional information, see

Pas et al., 2016). The CCU coaching was provided by four research team-trained and supervised

coaches who were external to the participating schools and had a background in

coaching/consultation. Each coach had either a master's in education or doctorate in school

psychology. On average, coaches logged 9.85 contacts with the teachers who were coached.

Coaches dedicated an average of 511.41 minutes (i.e., 8.52 hours) to each coaching case, of

which 221.25 minutes (i.e., 3.69 hours) required direct contact with the teacher. Specifically, no

time was required of teachers for coaches to complete observations and to prepare and plan for

coaching sessions. Additional data documenting the fidelity of the coaching (e.g., dosage) are

summarized in Pas et al. (2016); data also indicated that teachers and coaches viewed their

working relationship, investment in the coaching, the coaching process, and perceived benefits

very positively (see Johnson, Pas, & Bradshaw, 2016).

Measures

The sections below summarize the various sources of data and related measures collected

in the current study.

Office disciplinary referrals (primary outcome). Participating schools routinely

tracked ODRs made by teachers within the school using the School-Wide Information System

(SWIS; Irvin et al., 2004; May et al., 2003). For the current study, SWIS data were downloaded,

thereby providing the total number of ODR referrals (i.e., counts of specific incidents) made by

each teacher participating in the study for each school year, including the year prior to the

intervention (2012-13) and the year of the intervention (2013-14). The count data were available

for students overall and were disaggregated by student race/ethnicity (i.e., White and Black as

the two largest racial/ethnic groups), gender, and a combination of both (i.e., Black males).

Although prior research has demonstrated the validity of ODR data as an indicator of student

behavior problems (Pas, Bradshaw, & Mitchell, 2011), in the present study, we measured

referrals as an indicator of a complex set of interactions among a number of individuals (i.e.,

teachers, students, administrators) that may reflect a student's problem behavior, or differences

in behavioral and cultural standards and expectations, a teacher's lack of relational or behavior

management skills, or teacher bias (Irvin et al., 2004). Operationalizing *disproportionality* in

ODRs is an ongoing subject of debate with multiple decision points (e.g., characterizing differences in relative or absolute terms, benchmarking rates compared to White students or all students) and consensus is lacking on a single, optimal measure (Bottiani, Bradshaw, & Mendelson, 2017). Nonetheless, there is general agreement that, because of the potential harm associated with exclusionary discipline, it is critical to assess whether both gaps *and* incidence are reduced as a result of an intervention, rather than just examining reductions in gaps alone (Losen, Hodson, Keith et al., 2015). For example, one would not want gaps to be reduced by referrals of White students increasing while Black students' referrals remained the same; rather, the goal is to reduce disparities for Black students while also showing reductions or low rates of referrals overall. To assess disproportionality in a way that captured both relevant dimensions, we analyzed changes in referral incidence separately for each racial/ethnic group, with the rationale that significant reductions for Black students, in the absence of significant increases for White or other non-Black students, would indicate reductions in disproportionality as well.

Classroom observations (primary outcomes). External observers hired by the research team conducted classroom observations using the *Assessing School Settings: Interactions of Students and Teachers* (ASSIST; Rusby, Crowley, Sprague, & Biglan, 2011; Rusby, Taylor, & Milchak, 2001). Training in the ASSIST was provided in four stages: 1) an initial didactic session; 2) practice at non-project schools; 3) inter-observer agreement or reliability; and 4) on-site recalibration during data collection. During the training, observers were provided a manual detailing the ASSIST classroom procedures, including operational definitions of all codes, and step-by-step observational recording procedures. Following the initial training and practice sessions, observers completed the on-site inter-observer agreement or calibration session at non-project schools. Each observer trainee was paired with alternating expert observers in three

classrooms for simultaneous data coding for practice and reliability sessions. Calibration continued until a criterion of 80% was met; additional classroom observations were conducted if necessary to meet this criterion. Finally, on at least one occasion during active data collection, observers were joined by an expert observer for an on-site recalibration session. Average inter-observer agreement during recalibration was 87% for this study. For more information on the psychometric properties of the ASSIST, including its predictive validity, see Pas, Cash, O'Brennan, Debnam, and Bradshaw (2015); Debnam, Pas, Bottiani, Cash, and Bradshaw (2015); and Bottiani, Lindstrom Johnson, Debnam, and Bradshaw (2018).

The ASSIST was administered at each time-point through a single, 15-minute live classroom observation of a teacher providing instruction, during which the observers recorded event-based tallies; after the 15-minute live observation, the observer provided global ratings of student and teacher behaviors. Specifically, tallied teacher behaviors included the use of proactive behavioral management, opportunities to respond, approval, and disapproval in the classroom. Student non-cooperation and disruptions were also tallied. The total number of tallied behaviors for each behavior code was included as an outcome of this study. The global ratings included the following subscales: *teacher direction and influence* (5 items; α =.77; "There is evidence of classroom routines – students know what they're supposed to be doing"), *teacher anticipation and responsiveness* (6 items; α = .86; "Teacher anticipates when students may have problems behaviorally"), *teacher proactive behavior management* (4 items; α = .64; "Teacher gives clear instructions and directives to students"), *culturally responsive teaching strategies* scale (7 items; α = .73; "Teacher integrates cultural artifacts reflective of students' interests into learning activities"), *teacher and student meaningful participation* (8 items; α=.88; "Students are provided opportunities to contribute to discussion"), *student cooperation* (7 items; α = .92;

"Students handle transitions well") and *student socially disruptive behavior* (3 items; α = .51; "Students argue with peers"). Observers responded to 5-point Likert-type items ranging from 0 for *never* to 4 for being observed *almost continuously* for nearly all scales; observers responded on a scale of 0 for *never occurred* to 4 for *often occurred (6+ times)* for the student socially disruptive behaviors. These subscales were confirmed through factor analysis (CFI = .981, TLI = .979, RMSEA = .035, WRMR = 1.111). Scale scores were created by averaging the items on each subscale, with higher scores reflecting that the teacher or students engaged in these behaviors more frequently.

Teacher self-report surveys (secondary outcomes). Teachers were asked to complete online surveys regarding their perceived cultural responsivity, behavioral self-efficacy, and work-related stress on a 6-point Likert scale from 1 (*strongly disagree*) to 6 (*strongly agree*). We summarize these measures below. In all cases, scale scores which serve as outcome measures were constructed by averaging the responses on the items. Higher scores indicated more of the measured construct. Teachers also provided basic demographic data about themselves to be used as control variables in the outcome analyses, also described below.

Culturally responsive practices self-efficacy. The teacher survey included two different measures of self-efficacy to implement culturally responsive practices. The *Multicultural Efficacy Scale* (15 items; Guyton & Wesche, 2005; α = .80) assessed teachers' self-efficacy to provide culturally responsive instruction (e.g., "I can help students take on the perspective of ethnic and cultural groups different from their own;" "I can develop activities that increase the self-confidence of diverse students."). The *Culturally-Responsive Teaching Self-Efficacy Scale* (15 items; Siwatu, 2007; α = .82) also measured teachers' culturally responsive self-efficacy in addition to the teachers' assessment of their ability to connect with diverse students (e.g., "I

implement strategies to minimize the effects of mismatches between my students' home culture and the school culture" and "I use my students' cultural background to create a meaningful learning experience"). Higher scores were desirable, indicating higher efficacy.

Classroom behavior management self-efficacy. We also included an efficacy scale from Hoy and Woolfolk (1993) which assessed teachers' ability to handle students with behavior problems (e.g., "I can manage almost any student behavior problem"; $\alpha = .80$). This measure was selected to be specific to efficacy regarding the handling of student behavioral problems rather than general teaching efficacy given the nature of the intervention. Higher scores were desirable, indicating more efficacious teachers.

Stress. We utilized the work-related stress scale (Hurrell & McLaney, 1988), which was comprised of five items (e.g., "In my job, I feel I am under great stress"; "I am unable to cope with the stress of my job on a daily basis"; $\alpha = .84$). Higher scores were undesirable, indicating more teacher stress.

Social desirability bias. We utilized an abbreviated version of the social desirability bias scale (Crown & Marlowe, 1960), which was comprised of ten items (e.g., "I have never intensely disliked any of my students;" "I'm always willing to admit it when I make a mistake;" "I always try to practice what I preach"; $\alpha = .62$). A higher score on this scale indicated greater bias of responses due to social desirability. This variable was included as a covariate in the analyses of the self-report data to adjust for the potential of social desirability bias, which is a common concern in self-report measures of cultural responsivity (Larson & Bradshaw, 2017).

Demographics. Teachers reported their demographic characteristics, including race/ethnicity, age, and gender. For the purpose of this analysis, these variables were dichotomized, such that White was coded 1 and all other races were coded 0; early career was

coded 1 (age $\geq$30), whereas more experienced teachers were coded 0 (>31 years of age); and

female gender was coded 1, whereas male was coded 0. Schools were also characterized as either

middle (coded 1) or elementary (coded 0).

Analyses

The sections below summarize the analyses conducted to test the primary research

questions.

Multilevel analysis of pre-post changes leveraging the non-experimental design. We

fit hierarchical linear models (HLM; Raudenbush & Bryk, 2002; Raudenbush et al., 2011) to

examine the non-experimental difference in outcomes for the full sample ($N = 158$) from pre-test

(time 1/baseline/fall) to post-test (time 2/spring), without regard to coaching exposure. We

analyzed a three-level model with outcome (i.e., continuous observational measure and self-

reported survey scales and total counts on office disciplinary referrals and ASSIST tallies) and

time at level one, teacher characteristics at level two (White race, female gender, early career),

and school (i.e., middle vs. elementary) at level three. For teacher self-report of efficacy with

classroom skills (i.e., with regard to culturally responsive teaching, multicultural sensitivity, and

behavior management), we adjusted for social-desirability bias (Larson & Bradshaw, 2017). We

examined whether the slope of time was significant for each outcome and directionality, to

determine whether there was a significant increase or decline in each outcome variable.

Continuous variables were grand-mean centered to facilitate interpretation.

Multilevel analysis of coaching effects leveraging the experimental design. To

determine whether coached (versus non-coached) teachers demonstrated better scores on the

ODR, self-report, and ASSIST outcomes, we conducted two-level HLM analyses and accounted

for the nesting of teachers within schools as well as the school type (i.e., elementary versus

middle). At the individual teacher level (i.e., Level 1), the pre-test (i.e., baseline/fall) scale, teacher age, teacher gender, and teacher race were controlled for while intervention status was tested (i.e., because teachers were randomized within schools). At Level 2, the school level (0 = elementary versus 1= middle) was modeled and error arising from the non-independence of teachers within schools was appropriately accounted for. The significance and direction of intervention status coefficient were of specific interest here. Continuous variables were grand-mean centered to facilitate interpretation. Specifically, we modeled the fixed effects of each level-1 predictor including intervention status. Thus, the only random variance parameters included were for the intercept and level-1 error. Because our main focus was on the fixed main effect of the coaching intervention, we did not model the random effects. Further, with a limited school sample size of just 12 schools and thus limited power to detect cross-level interactions to explain the random effects of level-1 predictors, this was not a feasible area to explore in the current study.

Distributional assumptions. All scale scores (i.e., global ratings on the ASSIST and the teacher survey) were modeled with the normal distribution, as these scores were continuous; therefore, we report beta coefficients for these outcomes. In the case of the *socially disruptive student behavior* ASSIST scale and teacher self-report *stress* scale, a negative coefficient was desirable whereas a positive coefficient was desirable for all other ASSIST and teacher self-report scales. The tallied behaviors on the ASSIST and office disciplinary referrals were count variables and therefore the Poisson distribution was used. This distribution accounts for the fact that these outcome scores are bounded at 0 on the low end and are unbounded on the high end (Cameron & Trivedi, 1998). In instances where count outcomes were analyzed and the variance exceeded the mean, overdispersion was also accounted for, as suggested by Cameron and Trivedi

(1998). The Poisson regressions produced coefficients in log form, which were exponentiated to present more easily interpretable incidence rate ratios (IRR). In instances of ODRs and negatively-worded ASSIST tallies (i.e., disapproval, student non-cooperation, and student disruption), an incident rate ratio below 1 was desirable whereas an incident rate ratio above 1 was desirable for positively-worded ASSIST tallies (i.e., proactive behavior management, opportunities to respond, and approval).

Effect sizes. For the changes from pre- to post-test assessments and the coaching intervention effects, we calculated effect sizes for interpretability. For the pre-post changes, unadjusted means were used to calculate a Cohen's *d* (Cohen, 1992) where an effect of up to .20 was considered small, from 0.20 to 0.50 was considered moderate, and above 0.50 was considered large. For the coaching intervention effects, Spybrook's deltas (i.e., Δ) were calculated using adjusted HLM coefficients for the intervention effect, divided by the pooled standard deviation (Spybrook, 2008). The same interpretations of magnitude of effects apply.

Results

We first present the results for the primary outcomes of interests, which were largely the behavioral data. In this section, we also report the experimental findings from the randomized study design, in which we contrast the behavioral changes observed for the coached teachers in relation to the non-coached teachers. The second set of findings focus on behavioral changes observed the non-experimental pre-post contrasts, thereby testing changes between fall and spring of the school year.

Primary Outcomes of Interest

Office disciplinary referrals. No significant differences were identified in the non-experimental pre-post comparisons for ODRs (see Table 3). Comparing teachers randomly assigned to receive coaching to those who did not (see Table 4), coached teachers demonstrated

significantly improved IRR of referrals of Black students. Specifically, referrals for Black students (β = -0.28, IRR = 0.76, p = .015) were significantly lower among those who received coaching. On average, teachers who were coached had a predicted value of 1.43 referrals of Black students during the school year, as compared to the non-coached teacher predicted value of 1.99 (i.e., the fully unconditional incident rate). IRRs and differences in the average referrals of White students and students overall were non-significant.

Table 3. *Three-level repeated measures analysis within schools examining non-experimental pre-post change* (*N* = 158)

	Intercept	Time	White	Early Career	Female	Middle	
Disciplinary Referrals				Incident Rate Ratios (CI)			
Total referrals	2.36 (1.9, 3.0)	1.00 (0.8, 1.2)	2.38*** (1.5, 3.7)	1.00 (0.7, 1.4)	0.73 (0.5, 1.2)	2.09** (1.3, 3.3)	
Black students' referrals	1.43 (1.0, 2.0)	0.99 (0.8, 1.2)	2.14** (1.4, 3.4)	1.03 (0.7, 1.5)	0.77 (0.5, 1.3)	2.22* (1.2, 4.2)	
White students' referrals	0.38 (0.3, 0.6)	0.87 (0.7, 1.2)	3.18** (1.6, 6.4)	0.72 (0.4, 1.2)	0.66 0.4, 1.2)	1.66 (0.7, 4.2)	
ASSIST Tallies							
Proactives	7.71 (6.7, 8.9)	1.09 (1.0, 1.3)	1.05 (0.9, 1.3)	1.13 (0.9, 1.4)	0.99 (0.8, 1.3)	0.66* (0.5, 0.9)	
Approvals	6.37 (5.5, 7.4)	0.91 (0.7, 1.1)	1.00 (0.7, 1.3)	0.98 (0.8, 1.3)	1.19 (0.8, 1.8)	0.59** (0.5, 0.8)	
Disapprovals	0.18 (0.9, 1.3)	0.64** (0.5, 0.9)	0.91 (0.4, 1.9)	1.04 (0.5, 2.1)	0.99 (0.4, 2.5)	0.69 (0.3, 1.5)	
Opportunities to Respond	22.35 (20.4-24.4)	1.05 (0.9, 1.2)	0.85 (0.7, 1.0)	1.00 (0.9, 1.2)	1.01 (0.8, 1.3)	0.82* (0.7, 1.0)	
Non-cooperation	0.17 (0.1, 0.3)	2.04*** (1.5, 2.8)	1.36 (0.6, 3.1)	0.83 (0.4, 1.7)	0.99 (0.4, 2.5)	1.72 (0.6, 4.9)	
Disruptive Behaviors	10.59 (9.3, 12.0)	0.96 (0.8, 1.1)	1.36* (1.1, 1.7)	1.01 (0.8, 1.2)	1.05 (0.8, 1.4)	1.68** (1.3, 2.2)	
ASSIST Global Ratings				β coefficients (SE)			
Cultural Responsiveness	1.36 (0.09)	-0.16 (0.11)	-0.12 (0.14)	-0.24† (0.13)	0.25 (0.18)	-0.17 (0.19)	
Teacher Direction & Influence	3.58 (0.06)	-0.05 (0.05)	-0.05 (0.08)	-0.01 (0.08)	0.08 (0.11)	-0.19 (0.11)	
Teacher Responsiveness	3.23 (0.09)	-0.18* (0.09)	-0.04 (0.11)	0.03 (0.10)	0.12 (0.14)	-0.14 (0.17)	
Teacher Proactive	3.12 (0.07)	-0.22** (0.07)	-0.03 (0.10)	-0.01 (0.09)	0.26* (0.12)	-0.23 (0.14)	
Meaningful Participation	2.85 (0.13)	-0.15 (0.10)	-0.12 (0.13)	-0.10 (0.12)	0.30† (0.16)	-0.23 (0.26)	
Student Cooperation	3.35 (0.09)	-0.09 (0.11)	-0.03 (0.07)	0.03 (0.05)	0.19 (0.28)	-0.23 (0.16)	
Student Social Disruptions	0.78 (0.03)	-0.40*** (0.05)	0.07 (0.06)	-0.05 (0.06)	-0.07 (0.08)	0.03 (0.06)	
Teacher Self-Report	Intercept	Time	White	Early Career	Female	Middle	SDB±
CRT Efficacy	4.45 (0.06)	0.26*** (0.04)	-0.06 (0.09)	-0.04 (0.08)	-0.05 (0.11)	-0.05 (0.12)	0.34*** (0.06)
Multicultural Efficacy	4.66 (0.03)	0.15*** (0.05)	-0.08 (0.08)	-0.02 (0.07)	-0.17† (0.09)	0.06 (0.07)	0.28*** (0.05)
Behavioral Mgmt. Efficacy	4.54 (0.05)	0.18*** (0.05)	0.04 (0.11)	0.03 (0.10)	-0.20 (0.13)	0.03 (0.10)	0.47*** (0.07)
Stress	3.61 (0.11)	-0.14* (0.07)	0.41* (0.18)	-0.16 (0.16)	0.38 (0.22)	0.01 (0.21)	--

Note. Significant findings were denoted as * for *p* < .05; ** for *p* < .01; *** for *p* < .001. †denotes a marginally significant finding at *p* < .10. Tallies and referrals analyses assumed an over-dispersed Poisson distribution. CRT = Culturally responsive teaching; SDB = Social desirability bias. ±Teacher self-report adjusted for SDB.

Table 4. *Two-level analysis within schools examining main experimental effects of coaching* (*N* = 158)

	Intercept	Coached	Time 1	White	Early Career	Female	Middle
Disciplinary Referrals				Incident Rate Ratios (CI)			
Total referrals	1.58 (1.0, 2.5)	0.88 (0.7, 1.1)	1.04*** (1.0, 1.1)	1.94*** (1.5, 2.6)	1.28* (1.1, 1.5)	0.63 (0.5, 0.8)	1.69* (1.1, 2.6)
Black students' referrals	1.00 (0.5, 1.8)	0.76* (0.6, 0.9)	1.04*** (1.0, 1.1)	1.58** (1.2, 2.2)	1.47** (1.2, 1.9)	0.66** (0.5, 0.9)	2.16* (1.2, 3.9)
White students' referrals	0.18 (0.1, 0.6)	1.39 (0.9, 2.2)	0.97 (0.8, 1.1)	5.21*** (1.9, 14.3)	0.60† (0.3, 1.0)	0.52* (0.3, 0.9)	1.18 (0.6, 2.5)
ASSIST Tallies							
Proactives	7.62 (4.2, 13.8)	1.03 (0.9, 1.3)	1.02* (1.0, 1.0)	1.24 (1.0, 1.6)	1.07 (0.9, 1.3)	1.12 (0.8, 1.6)	0.69 (0.4, 1.2)
Approvals	4.30 (1.8, 10.4)	1.41† (1.1, 1.8)	1.00 (0.9, 1.3)	0.97 (0.6, 1.5)	1.02 (0.8, 1.4)	1.49 (0.8, 2.8)	0.68 (0.4, 1.1)
Disapprovals	0.18 (0.0, 0.9)	1.61 (0.7, 3.5)	1.00 (0.7, 1.4)	0.58 (0.3, 1.2)	2.09* (0.9, 1.3)	0.99 (1.0, 4.3)	0.17 (0.5, 2.8)
OTR	25.27 (15.8, 40.4)	0.98 (0.8, 1.2)	1.00 (1.0, 1.0)	0.82* (0.7, 1.0)	1.11* (0.9, 1.4)	1.15* (0.8, 1.7)	0.90 (0.7, 1.2)
Non-cooperation	0.57 (0.9, 3.5)	0.42* (0.3, 0.7)	1.34† (1.2, 1.5)	0.42 (0.3, 0.6)	0.78 (0.4, 1.5)	2.75 (1.0, 7.9)	1.69 (0.5, 6.1)
Disruptive Behaviors	5.49 (2.6, 11.5)	0.96 (0.8, 1.1)	1.01* (1.0, 1.0)	1.38* (1.1, 1.7)	1.14* (1.0, 1.4)	0.98 (0.7, 1.4)	2.12 (1.1, 4.0)
ASSIST Global Ratings				β coefficients (SE)			
Cultural Responsiveness	1.02 (0.32)	0.18 (0.15)	0.12 (0.7)	0.11 (0.18)	-0.22 (0.16)	0.07 (0.23)	0.08 (0.22)
Teacher Direction/Influ	3.52 (0.18)	0.13 (0.09)	0.39* (0.09)	-0.02 (0.10)	-0.08 (0.10)	0.04 (0.13)	-0.11 (0.09)
Teacher Responsiveness	3.01 (0.28)	0.29* (0.14)	0.12 (0.10)	0.05 (0.16)	-0.01 (0.15)	0.03 (0.21)	-0.19 (0.16)
Teacher Proactive	2.65 (0.33)	0.34* (0.13)	0.25* (0.16)	0.01 (0.09)	-0.13 (0.16)	0.33 (0.26)	-0.24 (0.16)
Meaningful Participation	2.24 (0.36)	0.14 (0.18)	0.14 (0.10)	0.18 (0.21)	-0.19 (0.19)	0.42 (0.26)	0.00 (0.22)
Student Cooperation	3.16 (0.43)	0.34* (0.15)	0.32* (0.11)	0.11 (0.12)	-0.18 (0.07)	0.07 (0.27)	-0.26 (0.14)
Student Disruptions	0.61 (0.16)	-0.19* (0.08)	0.02 (0.08)	0.10 (0.09)	-0.03 (0.08)	-0.12 (0.11)	0.21 (0.10)

Teacher Self-Report	Intercept	Coached	Time 1	White	Early Career	Female	Middle	SDB±
CRT Efficacy	4.57 (0.04)	-0.02 (0.08)	0.58*** (0.06)	-0.09 (0.09)	-0.00 (0.08)	-0.08 (0.11)	-0.12 (0.08)	0.26*** (0.06)
Multicultural Efficacy	4.76 (0.02)	0.001 (0.07)	0.52*** (0.08)	-0.00 (0.07)	-0.05 (0.06)	-0.04 (0.09)	-0.07 (0.06)	0.22*** (0.06)
Beh. Mgmt. Efficacy	4.64 (0.05)	0.09 (0.10)	0.52*** (0.07)	0.01 (0.11)	0.05 (0.10)	-0.16 (0.13)	-0.02 (0.10)	0.26** (0.08)
Stress	3.54 (0.08)	-0.20 (0.13)	0.70*** (0.06)	-0.00 (0.15)	0.19 (0.13)	-0.07 (0.18)	0.27 (0.16)	--

Note. Significant findings were denoted as * for *p* < .05; ** for *p* < .01; *** for *p* < .001. †denotes a marginally significant finding at *p* < .10. Tallies and referrals analyses assumed an over-dispersed Poisson distribution. CRT = Culturally responsive teaching; SDB = Social desirability bias. OTR = Opportunities to Respond. ±Teacher self-report adjusted for SDB.

ASSIST observations. One significant difference in the non-experimental pre-post comparisons for ASSIST tallies was found for teachers' reduced use of *disapprovals* (IRR = 0.64, p = .007; Cohen's d = -0.15 see Table 3). In these non-experimental pre-post comparison analyses, we also found less teacher use of *proactive behavior management* (β = -0.22, p = .002; d = -0.28) and *anticipation and responsiveness* (β = -0.18, p = .04; d = -0.19) on the ASSIST global ratings.

When comparing those teachers randomly assigned to receive coaching and those who served as comparison (see Table 4), coached teachers were rated by outside observers as employing significantly better classroom management. Specifically, observer ratings of *proactive behavior management* (β = 0.34, p = .01; Δ = 0.45) and *anticipation of and responsiveness* to student problems (β = 0.29, p = .04; Δ = 0.37) were higher among those who received coaching. In other words, although all teachers were observed to improve on these classroom management measures, coached teachers were rated nearly one-half and over one-third of a standard deviation higher by observers on their use of proactive behavior management and their ability to anticipate and respond to student problems. Further, there was a marginally significant trend for teacher use of *approvals*, such that use of approvals demonstrated a slight increase (β = 0.34, IRR = 1.41, p = .09; Δ = 0.05) for coached teachers. On average, teachers who were coached used 10.19 approvals during the 15-minute observation, as compared to the overall average of 7.23 (i.e., fully unconditional IRR).

Analysis of the ASSIST student data also demonstrated significant improvements for the coached classrooms, relative to the non-coached classrooms; however, there were no significant effects on the ASSIST student data in the non-experimental pre-post comparisons. Specifically, regarding the experimental design contrast, observers documented more favorable ratings of

DOUBLE CHECK   27

student cooperation ($\beta = 0.34$, $p = .03$; $\Delta = 0.45$) in coached classrooms relative to non-coached

classrooms. With regard to the tallied student behaviors, the average number of times *non-*

cooperation was observed within non-coached classrooms was 0.31 (i.e., fully unconditional

IRR); for coached teachers, the average was 0.13 ($\beta = -0.86$, IRR = 0.42 $p = .002$; $\Delta = -0.48$).

Observers also recorded less *socially disruptive behaviors* ($\beta = -0.19$, $p = .04$; $\Delta = -0.43$).

Students in classrooms of coached teachers demonstrated nearly one-half of a standard deviation

greater student compliance and less socially disruptive behaviors.

Secondary Outcomes of Interest

In this section we report the secondary outcomes of interest, which were largely the self-

reported perceptual data. Specifically, we report the experimental findings from the randomized

study design, in which we contrast the self-reported data for the coached teachers in relation to

the non-coached teachers. The second set of findings focus on changes in the self-reported data

in the non-experimental pre-post contrasts, thereby testing changes between fall and spring of the

school year.

Teacher self-reported ratings. In the non-experimental pre-post analyses, we found

significant improvements at post-test on culturally responsive teaching self-efficacy ($\beta = 0.26$, p

$< .001$; $d = 0.48$), multicultural efficacy ($\beta = 0.15$, $p < .001$; $d = 0.32$), classroom behavior

management self-efficacy ($\beta = 0.18$, $p < .001$; $d = 0.26$), and stress ($\beta = -0.14$, $p = .037$; $d = -$

0.15) (see Table 3). In other words, all teachers rated their efficacy higher and stress lower at the

end of the school year. However, none of the teacher-reported survey scales demonstrated

significant differences between the coached and non-coached teachers in the experimental

contrast (see Table 4). In other words, the coaching did not boost the overall gains on any of the

self-reported efficacy or stress scales when comparing coached versus non-coached teachers.

Adjustment for Multiple Contrasts

We calculated the false discovery rate to ensure that the number of tests did not account for our findings. When set to 0.1 (i.e., 10%), $p < .03$ was significant (i.e., > 0.03 was not). Considering this lower p-value resulted in the loss of 2 out of the 12 originally-reported significant findings. Specifically, the pre-post non-experimental finding for teacher stress and the experimental finding of improvements for coached teachers on the global ASSIST anticipation scale were no longer significant. At the more conservative level of 0.05 (5%), we lost one additional result (i.e., the experimental finding that coached teachers have a lower rate of ODRs for Black students). Together, these false discovery rate adjustments suggested that a majority of the significant effects reported above are not likely due to the number of tests conducted/number of outcomes assessed.

Discussion

Despite persistent concern regarding the gaps in both discipline and academic performance between Black students and their peers, relatively few models have been developed or rigorously tested to determine their impact on these gaps or related outcomes (Bottiani et al., 2017). The current study examined the impact of the Double Check coaching, over and above the 5-part professional development series and SW-PBIS activities. Using a teacher-level RCT (experimental) design in 12 elementary and middle schools, we found evidence of a significant value added associated with the Double Check coaching on our primary behavioral outcomes of interest, including the use of ODRs and observed teacher practices and student behaviors. Specifically, the experimental contrasts indicated that decreases in the use of ODRs by coached teachers were most salient for ODRs of Black students, relative to non-coached teachers; importantly, the effect sizes for the ODR findings were relatively small and may potentially be

sensitive to the number of tests conducted. Similarly, the external observations indicated improvements in coached teachers' use of proactive behavior management and anticipation and responsiveness to student needs, relative to comparison teachers. This finding is consistent with the hypothesized impact of the coaching on these proximal classroom management practices. The experimental design contrasts also indicated that student cooperative and disruptive behaviors improved significantly, as assessed by external observations. These experimental findings leveraged the RCT design and allowed for stronger conclusions regarding the causal impact of coaching on external observations of student and teacher behavior. It is also important to note, however, that there were no significant pre-post differences in ODRs among the full sample (non-experimental). Moreover, only minimal changes in teacher practices were detected on the ASSIST in non-experimental pre-post contrasts (e.g., slightly lower rates of positive and proactive behavior management at the spring assessment relative to the fall assessment).

With regard to our secondary outcomes of interest, which were assessed through teacher self-reports, we detected pre-post non-experimental changes in teachers' self-reported culturally responsive behavior management and stress; however, there was no differential experimental effect of the coaching on these self-reports. These findings were robust to other teacher-level covariates and correlates, like experience, gender, or grade level (i.e., middle vs. elementary), as well as self-reported social desirability.

Taken together, the significant effects on the self-reported efficacy and stress outcomes for the pre-post non-experimental contrast, but not for the experimental contrasts, suggest that the professional development may have resulted in improvements in teacher beliefs, efficacy, and self-assessments of culturally responsive behavior management as well as reductions in stress. However, the coaching did not provide further benefit for these outcomes. In contrast, the

experimental comparisons indicated significant effects of the coaching on the behavioral data, including ODRs and external observations of teacher practices and student behaviors. This pattern of findings is consistent with prior research which suggests that professional development provided in a workshop format may improve knowledge, attitudes, and beliefs, but is less effective at improving teacher and student behaviors (Joyce & Showers, 2002). As such, one potential conclusion from the current pattern of findings is that the professional development may help improve the attitudinal self-report outcomes, but not affect the behavioral outcomes. In contrast, the coaching may be improving the behavioral outcomes, but not the attitudinal outcomes. In fact, the reason that Double Check was designed to include a traditional group-based professional development format supplemented with individual coaching was to leverage the efficiency of workshops, and augment it with the on-going support of coaching in order to effect behavioral changes in the classroom.

Although not a central focus of the study, we also observed a few trends in relation to the covariates. For example, the non-experimental pre-post contrasts suggested that middle school teachers improved the most on the ASSIST observations over the course of the year. This may have been because middle school teachers had greater room for improvement, as reflected in their poorer scores than elementary school teachers on all measures (e.g., less use of positive classroom management strategies, more student misbehavior). In contrast, elementary school teachers appeared to experience greater improvements on the self-reported teacher survey data. These trends should be explored further in larger and more diverse samples of schools, both in isolation and in relation to interactions with intervention status; the current study was not sufficiently powered to explore the effects.

Limitations and Future Directions

It is important to consider some limitations when interpreting these findings. For example, we focused on schools that had prior training in SW-PBIS (i.e., tier 1); this element was augmented through coaching and supports related to the review and use of ODR and other discipline data. As a result, it is not possible to determine from the current study design whether these augmentations or the base tier 1 SW-PBIS supports systematically impacted the pattern of findings. In fact, prior research on SW-PBIS suggests it alone does not narrow discipline gaps per se (Bradshaw, Waasdorp, & Leaf, 2012; Vincent et al., 2011; Vincent, Sprague, Pavel, Tobin, & Gau, 2015), but holds promise as a foundation for the augmented Double Check SW-PBIS activities. Similarly, all school staff were exposed to the five-part Double Check professional development model, and thus only the pre-post comparisons were relevant for exploring the possible effects of this element of the multi-component Double Check model. The current study and the RCT design focused on teachers; however, other non-teaching staff (e.g., psychologists, special educators) attended the professional development sessions and may have experienced positive impacts of the program that were not measured in the current study.

As with many multi-component programs, it is often difficult to isolate the effect of a particular element of a model. Nevertheless, it is important to recognize that all three elements of the Double Check model were not equally weighted; by design, coaches spent the majority of their time coaching teachers (51% of coach time logged), followed by implementing the professional development (11% of coach time logged), and then SW-PBIS team focused activities (< 5% of coach time logged). Given the relative investment in coaching, we specifically isolated the effect of the coaching in the RCT design in order to examine the effects of coaching, above and beyond the other two elements. Additional research is needed to determine the contributions of the other two elements (i.e., SW-PBIS augmentations and

professional development series), as well as on the potential synergist effects of the full Double

Check model. A separate RCT is currently being conducted of the full Double Check model that

includes school-level randomization; this ongoing study will allow us to better determine

whether the entire multi-component Double Check model improves all outcomes (i.e., attitudes,

beliefs, *and* behaviors) as compared to a business as usual condition.

The inclusion of a broad set of outcomes is both a strength and a limitation of this study.

We conceptualized the outcomes into primary (i.e., behavioral) and secondary outcomes (i.e.,

teacher self-reported attitudes and perceptions). Both sets of outcomes are articulated in the logic

model (Figure 1). With regard to the secondary outcome, we not only wanted to see whether

these improved, but wanted to ensure that no iatrogenic effects emerged particularly on factors

such as stress or self-efficacy, given that discussing issues of culture, equity, and

disproportionality can often result in tension for school personnel (Tatum, 2017). Importantly, no

iatrogenic effects were observed on any of the outcomes. On the other hand, the relatively large

number of outcomes examined could increase the possibility of a type 1 error. Therefore, we

calculated the false discovery rate to ensure that the number of tests did not account for our

findings. As noted above, the false discovery rate adjustments suggested that the majority of the

effects would still hold, even after adjusting for the number of tests conducted/number of

outcomes assessed.

Given the multiple sources of data, including ODRs, teacher self-report data, and external

observations of students and teachers, there were some areas of divergence in the pattern of

findings across the measures. This suggests a need for additional research on valid and reliable

measures of culturally responsive behavior management and teaching within a larger sample. For

example, a few of the measures had slightly lower internal consistency estimates (i.e., alphas),

which may be a function of the relatively small sample size. The tally of non-cooperation had a low base rate occurrence which may have led to some inconsistency in findings related to this outcome. Additional measurement research is needed to better understand potential areas of congruence and divergence between these and other sources of data (Bottiani et al., 2018). Furthermore, student perceptions of the classroom environment would enhance our understanding of effects of coaching in culturally responsive behavior management.

Another measurement limitation was that disproportionality was only assessed with regard to referrals. The ASSIST classroom observations of student behavior represent all students in the classroom and cannot be examined specifically for any racial/ethnic group. Further, the referral data were lacking the disaggregated number of Black and White students in a classroom and thus risk ratios could not be calculated. To discern whether change over time in referral counts for a teacher was influenced by change in the number of Black and White students in their classroom, sensitivity analyses were conducted utilizing an ASSIST indicator of the number of non-White students in the classroom (i.e., observers counted the number of non-White students in the classroom at the time of the observation). These sensitivity analyses indicated that our findings remained substantively unchanged in terms of coefficients, IRRs, and significance.

We examined the effects of the Double Check coaching across a single school year. Additional research is needed to determine whether these effects are sustained into subsequent school years and generalize into other samples. Although the sample of students included a relatively high proportion of Black students, there were considerably fewer Hispanic students and other ethnic minority or culturally diverse students, precluding the examination of impacts for these other student groups. Further, with small samples of other racial/ethnic groups and

lacking data on the concentration of different student groups within the classroom, we could not examine whether this variable modified outcomes. For example, teachers with a high(er) proportion of non-White students in their classroom may feel more motivated to use culturally responsive approaches, than teachers in classrooms with all White students. Larger and more diverse samples of students, as well as staff, would allow for further exploration. Although the proportion of Black and non-Black students in the schools was not equal, this is not uncommon in schools in the U.S. (Tatum, 2017), particularly in this district where White students are the majority.

Given our relatively small sample size, we decided to focus largely on main effects models; however, we intend to examine more directly potential effect modifiers such as elementary vs. middle, student race/ethnicity, classroom context, or teacher race/ethnicity in future studies (see Bradshaw, Mitchell, O'Brennan, et al,. 2010). Similarly, mediators, which are articulated in the Double Check logic model (Figure 1) were not tested, but will be explored in future studies with larger samples.

Conclusions and Implications

Taken together, the findings from this teacher-level RCT indicated promising effects of the Double Check adapted version of the CCU as a value added over the school-wide professional development and PBIS activities, both on teacher practices and observed student behaviors. This value added finding is particularly promising, given the relatively limited additional teacher time associated with participation in the coaching (i.e., an average of just 3.69 hours of direct contact with the coach for teachers, and 8.52 coach hours per teacher); this suggests a rather impressive high 'return on investment', given the effect sizes (which ranged from .15 to .48) observed on the primary (i.e., behavioral) outcomes associated with coaching.

Among a dearth of evidence-based interventions, these results suggest that the Double Check model is one of only a few systematic approaches to promoting culturally responsive behavior management which has been rigorously tested *and* shown to be associated with improvements in either student or staff outcomes (Bottiani et al., 2017). Additional research is needed to determine the effects of the entire Double Check model, as well as the extent to which these findings generalize to other school levels, like high schools, or extend beyond this particular school district.

References

Bottiani, J. H., Bradshaw, C. P., Rosenberg, M. S., Hershfeldt, P. A., Pell, K. L., & Debnam, K. J. (2012). Applying Double Check to response to intervention: Culturally responsive practices for learning disabilities. Insight on learning disabilities: *Prevailing Theories to Validated Practices, 9(1),* 93-107.

Bottiani, J.H., Larson, K. Debnam, K., Bischoff, C., & Bradshaw, C.P. (2017). Promoting educators' use of culturally responsive practices: A systematic review of in-service interventions. *Journal of Teacher Education*, 1–19. doi:10.1177/0022487117722553

Bottiani**,** J.H., Lindstrom Johnson, S., Debnam, K.J., & Bradshaw, C. (2018). *Triangulating school climate: Areas of convergence and divergence in multilevel perspective.* Manuscript in preparation.

Bottiani**,** J.H., Bradshaw, C.P., & Mendelson, T.M. (2017). A multilevel examination of racial disparities in high school discipline: Black and White adolescents' perceived equity, school belonging, and adjustment problems. *Journal of Educational Psychology, 109, 532-545.* doi:10.1037/edu0000155

Bradshaw, C. P., & Rosenberg, M.S. (Eds). (2018). *Double Check: Promoting Culturally Responsive Behavior Management and Enhanced Student Engagement.* Guilford Press. Book in preparation.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of School-Wide Positive Behavioral Interventions and Supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12,* 133-148. doi:10.1177/1098300709334798

Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology, 102*(2), 508-520

Bradshaw, C. P., Waasdorp, T. E. & Leaf, P. J. (2012). Effects of School-Wide Positive Behavioral Interventions and Supports on child behavior problems. *Pediatrics, 130*(5), e1136-e1145. doi:10.1542/peds.2012-0243

Bradshaw**,** C.P., Koth, C.W., Bevans, K.B., Ialongo, N., & Leaf, P.J. (2008). The impact of school-wide Positive Behavioral Interventions and Supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*(4), 462-473. doi:10.1037/a0012883

Bradshaw, C.P., O'Brennan, L.M. & McNeely, C. A. (2008). Core competencies and the prevention of school failure and early school leaving. In N. G. Guerra, & C. P. Bradshaw (Guest Editors), Core competencies to prevent problem behaviors and promote positive youth development. *New Directions for Child and Adolescent Development, 122,* 19-32. doi:10.1002/cd.226

Bradshaw, C.P., Waasdorp, T.E., & Leaf, P.J. (2015). Examining variation in the impact of School-wide Positive Behavioral Interventions and Supports: Findings from a randomized controlled effectiveness trial. *Journal of Educational Psychology, 107*(2), 546-557.

Cameron, A. C. & Trivedi, P. K. (1998). *Regression Analysis of Count Data*. New York: Cambridge Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of

      psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Day-Vines, N., & Day-Hairston, B. (2005). Culturally congruent strategies for addressing the

      behavioral needs of urban, African American male adolescents. *Professional School*

      Counseling, *8*(3), 236-243.

Debnam, K. J., Pas, E. T., Bottiani, J., Cash, A. H., & Bradshaw, C. P. (2015). An examination

      of the association between observed and self-reported culturally proficient teaching

      practices. *Psychology in the Schools*, *52*(6), 533-548.

Debnam, K., Bottiani, J., & Bradshaw, C.P. (2017). Promoting culturally responsive practice to

      reduce disparities in school discipline among African American students. In N. M.

      Finigan-Carr (Ed.). *Linking Health and Education for African American Students'*

      Success. pp. 97-114. New York: Routledge.

Debnam, K., J., Lindstrom Johnson, S., Waasdorp, T. E., & Bradshaw, C. P. (2014). Equity,

      connection, and engagement in the school context to promote positive youth

      development. *Journal of Research on Adolescence, 24,* 447-459.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P. III, & Booth E.

      A. (2011). *Breaking schools' rules: A statewide study of how school discipline relates to*

      students' success and juvenile justice involvement. College Station, TX: Council of State

      Governments Justice Center, Public Policy Research Institute of Texas, A&M University.

      Retrieved from

      http://justicecenter.csg.org/files/Breaking_Schools_Rules_Report_Final.pdf

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. New York, NY:

      Teachers College Press.

Guyton, E. M., & Wesche, M. V. (2005). The multicultural efficacy scale: Development, item

　　　selection, and reliability. *Multicultural Perspectives, 7(4)*, 21-29.

Hershfeldt, P. A., Sechrest, R., Pell, K. L., Rosenberg, M. S., Bradshaw, C.P., & Leaf, P.J.

　　　(2009). Double-Check: A framework of cultural responsiveness applied to classroom

　　　behavior. *TEACHING Exceptional Children Plus, 6*(2), 2-18.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J.

　　　(2009). A randomized, wait-list controlled effectiveness trial assessing school-wide

　　　Positive Behavior Support in elementary schools. . *Journal of Positive Behavior*

　　　Interventions, 11, 133-144.

Horner, R., Todd, A., Lewis-Palmer, T., Irvin, L., Sugai, G., & Boland, J. (2004). The school-

　　　wide evaluation tool (SET): A research instrument for assessing school-wide positive

　　　behavior support. *Journal of Positive Behavior Interventions, 6*(1), 3-12.

Hoy, W. K. & A. E. Woolfolk (1993). Teachers' sense of efficacy and the organizational health

　　　of schools. *The Elementary School Journal, 93*(4), 355-372.

Hurd, N., Hussain, S.B., & Bradshaw, C. P. (2015). School disorder, school connectedness, and

　　　psychosocial outcomes: Moderation by a supportive figure in the school. *Youth &*

　　　Society, 1-23. doi: 10.1177/0044118X15598029

Hurrell, J.J. Jr. & McLaney, M.A. (1988). Exposure to job stress - A new psychometric

　　　instrument. *Scandinavian Journal of Work, Environment, and Health, 14,* 27-28.

Irvin, L., Tobin, T., Sprague, J., Sugai, G., & Vincent, C. (2004). Validity of office discipline

　　　referral measures as indices of school-wide behavioral status and effects of school-wide

　　　behavioral interventions. *Journal of Positive Behavior Interventions, 6*(3), 131-147.

Jennings, P. A., Frank, J. L., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2013).

    Improving classroom learning environments by Cultivating Awareness and Resilience in

    Education (CARE): Results of a randomized controlled trial. *School Psychology*

    Quarterly, 28, 374–390. doi: 10.1037/spq0000035

Jennings, P.A. (2015). *Mindfulness for teachers: Simple skills for peace and productivity in the*

    classroom. New York: Norton.

Johnson, S. R., Pas, E. T., & Bradshaw, C. P. (2016). Understanding and measuring coach-

    teacher alliance: A glimpse inside the 'black box'. *Prevention Science, 17*, 439-449. doi:

    10.1007/s11121-016-0633-8.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development*. 3rd ed.

    Association for Supervision and Curriculum Development; Alexandria, VA.

Larson, K. & Bradshaw, C. P. (2017). Cultural competence and social desirability among

    practitioners: A systematic review of the literature. *Children & Youth Services Review,*

    76, 100-111. DOI: 10.1016/j.childyouth.2017.02.034

Losen, D., Hodson, C., Keith, M. A., Morrison, K., & Belway, S. (2015). *Are we closing the*

    school discipline gap? Los Angeles, CA: The Center for Civil Rights Remedies at the

    Civil Rights Project, University of California Los Angeles. Retrieved from

    http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-

    remedies/school-to-prison-folder/federal-reports/are-we-closing-the-school-discipline-

    gap/AreWeClosingTheSchoolDisciplineGap_FINAL221.pdf

Losen, D.J., Hodson, C., Ee, J., & Martinez, T. (2015). Disturbing inequities: Exploring the

    relationship between racial disparities in special education identification and discipline.

In Losen, D.J., (Ed). *Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion* (pp. 89-106). New York: Teachers College Press.

May, S., Ard, W., Todd, A., Horner, R., Glasgow, A., Sugai, G., et al. (2003). Schoolwide Information System. Eugene: University of Oregon, Educational & Community Supports.

Miller, W. R., & Rollnick, S. (2002). *Motivational Interviewing: Preparing people for change*. (2nd ed.). New York: Guilford Press.

O'Connell, M. E., Boat, T., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions; Institute of Medicine; National Research Council. The National Academies Press.

Pas, E. T., Cash, A. H., O'Brennan, L., Debnam, K. J., & Bradshaw, C. P. (2015). Profiles of classroom behavior in high schools: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology*, *53*(2), 137-148.

Pas, E. T., Larson, K., Reinke, W., Herman, K., & Bradshaw, C.P. (2016). Implementation and acceptability of an adapted Classroom Check-Up coaching model to promote culturally responsive classroom management. *Education and Treatment of Children*, *39*(4), 467-491.

Pas, E., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools, 48*(6), 541-555. doi:10.1002/pits.20577

Porowski, A., O'Conner, R., & Passa, A. (2014). *Disproportionality in school discipline: An assessment of trends in Maryland, 2009-12*. Regional Educational Laboratory Mid-Atlantic, U.S. Department of Education.

Raudenbush, S. W. & A. S. Bryk (2002). *Hierarchical linear Models: Applications and data analysis methods.* Thousand Oaks, CA, Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. K., Fai, Y. F., Congdon, R. T., Jr., & duToit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Reinke, W. M., (2006). The Classroom Check-Up: Advanced tools for effective coaching. In Sprick, R., Knight, J., Reinke, W.M., & McKale, T. (Eds.). *Coaching for Positive Classrooms: Supporting Teachers with Classroom Management* (139-167). Eugene, OR: Pacific Northwest Publishing.

Reinke, W. M., Herman, K. C., & Sprick, R. (2011). *Motivational interviewing for effective classroom management: The Classroom Check-up*. New York: Guilford Press.

Reinke, W.M., Lewis-Palmer, T., & Merrell, K. (2008). The Classroom Check-up: A classwide consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37,* 315-332.

Richards, H., Brown, A., & Forde, T. (2007). Addressing diversity in schools: Culturally responsive pedagogy. *Teaching Exceptional Children, 39*, 64-68.

Ross, D., Kamman, M., & Coady, M. (2008). Accepting responsibility for the learning of all students. In M. Rosenberg, D. Westing, & J. McLeskey (Eds.), *Special education for today's teachers* (pp. 52-81). Upper Saddle River, NJ: Pearson.

Rusby, J. C., Taylor, T., & Milchak, C. (2001). *Assessing school settings: Interactions of students and teachers (ASSIST) observation system*. Unpublished manual. Eugene, OR: Oregon Research Institute.

Rusby, J.C., Crowley, R., Sprague, J., & Biglan, A. (2011). Observations of the middle school environment: Context for student behavior beyond the classroom. *Psychology in the Schools, 48*, 400-415.

Siwatu, K.O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, *23*, 1086-1101.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, *51*(4), 640-670.

Skiba, R. J., Horner, R. H., Chung, C., Rausch, M., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, *40*, 85-107.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review*, *34, 317-342;*

Skiba, R. J., Shure, L., & Williams, N. (2012). Racial and ethnic disproportionality in suspension and expulsion. In A. L. Noltemeyer C. S. Mcloughlin (Eds.), *Disproportionality in education and special education* (pp. 89-118). Springfield, IL: Charles C. Thomas Publisher, Ltd.

Spybrook, J. (2008). *Power, sample size, and design*. In A. A. O'Connell & D. B. McCoach

    (Eds.), Multilevel modeling of educational data (pp. 273-311). Charlotte, NC:

    Information Age Publishing.

Sugai, G. & Horner, R. (2002). The evolution of discipline practices: School-wide positive

    behavior supports. *Child & Family Behavior Therapy*, *24,* 23-50.

Sugai, G., & Horner, R. (2006). A promising approach for expanding and sustaining school-wide

    positive behavior support. *School Psychology Review, 35*, 245-259.

Tatum, B. D. (2017). *Why are all the Black kids sitting together in the cafeteria?: And other

    conversations about race*. Basic Books.

U.S. Department of Education, Office of Civil Rights, (2014, March). Civil Rights Data

    Collection Data Snapshot: School Discipline, Issue Brief No. 1. Retrieved from

    http://ocrdata.ed.gov/Downloads/CRDC-School-Discipline-Snapshot.pdf

Vincent, C. G., Randall, C., Cartledge, G., Tobin, T. J., & Swain-Bradway, J. (2011). Toward a

    conceptual integration of cultural responsiveness and schoolwide positive behavior

    support. *Journal of Positive Behavior Interventions*, *13*(4), 219-229.

Vincent, C. G., Sprague, J. R., Pavel, M., Tobin, T. J., & Gau, J. M. (2015). Effectiveness of

    schoolwide Positive Behavior Interventions and Supports in reducing racially inequitable

    disciplinary exclusion. Conference proceedings. *Closing the school discipline gap:

    Equitable remedies for excessive exclusion*, 207-221.

Wang, C. S., Tai, K., Ku, G., & Galinsky, A. D. (2014). Perspective-taking increases willingness

    to engage in intergroup contact. *PloS One, 9,* e85681. doi: 10.1371/journal.pone.0085681

Author Biographical Statements

Catherine P. Bradshaw, Ph.D., M.Ed., is a professor and the Associate Dean for Research and Faculty Development at the Curry School of Education at the University of Virginia. Her primary research interests focus on the development of aggressive behavior and school-based prevention. She has led a number of federally funded randomized trials of school-based prevention programs, including Positive Behavioral Interventions and Supports (PBIS) and social-emotional learning curricula. She has coauthored over 225 peer-reviewed publications and chapters on issues related to school-based prevention programming; she is the is the editor of the journal *Prevention Science*.

Elise T. Pas, PhD, is an Associate Scientist in the Department of Mental Health at the Johns Hopkins University, Bloomberg School of Public Health. Her research examines the effectiveness of evidence- based interventions delivered through a multi-tiered system of supports, coaching interventions, the influence of teacher and school contextual factors on implementation and student outcomes, and the adoption and implementation of positive behavior supports in schools. She holds a doctorate in school psychology from the University of Maryland, College Park.

Jessika H. Bottiani, Ph.D., M.P.H. is an assistant professor on the research faculty at the Curry School of Education at the University of Virginia. Her primary research interests center on marginalized students' inequitable experiences of safety, support, and belonging (i.e., school climate) in U.S. K-12 public schools. Her research examines the structural causes and consequences of inequitable school climate to inform the development and testing of school-based preventive intervention and promotion strategies to reduce racial disparities in school discipline.

Katrina Debnam, Ph.D., M.P.H. is an assistant professor in the School of Nursing and the Curry School of Education at the University of Virginia. Her primary research interests center on health disparities research, youth violence prevention, school-based prevention programming, and teen dating violence. Prior to her current appointment, she was on the faculty in the Department of Mental Health at the Johns Hopkins University, Bloomberg School of Public Health. She holds a doctorate in public health from the University of Maryland.

Wendy M. Reinke, Ph.D. is a Professor in School Psychology at the University of Missouri and Co-director of the Missouri Prevention Center. Her primary research interests are in training and supporting school personnel to deliver effective practices with a particular interest in personnel use of evidence-based social behavioral and emotional interventions. She is the co-author of seven books, including Motivational Interviewing for Effective Classroom Management: The Classroom Check-up, and has ten chapters and over 85 peer-reviewed publications related to prevention of social emotional and behavior problems.

Keith C. Herman, Ph.D. is a Professor in School and Counseling Psychology at the University of Missouri and Co-director of the Missouri Prevention Center. His primary research interests are in using Motivational Interviewing to increase engagement and effective practices in the classroom and supporting youth mental health. He has authored over 90 peer-reviewed publications and four books on behavior management, prevention, and early intervention.

Michael S. Rosenberg, Ph.D. is the dean of the School of Education at the State University of New York at New Paltz. As the Dean of the School of Education at New Paltz, Rosenberg serves as the chief academic and administrative officer for the School. He previously served as the associate dean for research in the School of Education and a professor in the Department of Special Education at Johns Hopkins University in Baltimore, MD.

Figure 1. Logic Model for Double Check Program

Inputs	Activities	Immediate Outcomes	Intermediate Outcomes	Ultimate Outcomes
Knowledge Base • Research on culturally responsive practices • Research on recommended strategies for promoting cultural responsivity • Evidence-based classroom management & student engagement strategies	**SW-PBIS** • Monthly meetings of PBIS team to review data • Coach provides support and assistance to PBIS team and administrator in data-based decision-making	**SW-PBIS** • Enhanced awareness of problem of disproportionality at school • Enhanced knowledge of strategies for using data to make decisions	**SW-PBIS** • Increased use of data to make decisions	**Student Discipline** • Reduction in discipline problems (e.g., office discipline referrals [ODRs], suspensions) for all students
SW-PBIS • Implementation of core elements of SW-PBIS	**Double Check Professional Development Series** • Provide staff with information on disproportionality • Conduct 5 part training on cultural responsivity and student engagement consistent with the 5 Double Check CARES domains • Promote dialogue about culture and equity among staff	**Cultural Responsivity** • Enhanced awareness of research on discipline disproportionality affecting students of color • Enhanced reflection on own cultural context • Increased comfort discussing issues of culture • Increased awareness of students' cultural context • Increased awareness of cultural and contextual setting events uniquely affecting students of color	**Cultural Responsivity** • Increased use of culturally appropriate discipline & engagement strategies • Increased formation of authentic relationship with students • Increased connection of curriculum to students' culture • Increased effective communication with students of color	**Student Achievement** • Improvement in academic performance for all students
Buy-In & Partnerships • District administration • Principals • Staff • Community & parents • State Dept of Education	**Coaching of Teachers Using the Double Check Adapted Version of the Classroom Check-Up (CCU)** • Conduct CCU motivational interview • Collect data on classroom ecology, student behavior, and teacher's classroom management • Conduct feedback session on data collected • Select from menu of evidence-based culturally sensitive interventions • Provide training in selected strategies • Provide ongoing performance feedback • Train teacher to self-monitor use of effective classroom management and student engagement strategies	**Enhanced Classroom Management** • Increased use of evidence-based classroom management and student engagement strategies	**Student Engagement** • Increased student engagement in academic curriculum • Increase in connectedness to school and teachers • Improved perceptions of school climate	**Reduction in Disproportionality** • Reduction in relative risk ratio for special education, suspensions, ODRs, school failure, dropout, and gifted and talented referral • Narrowing of the achievement gap
Time Allocation • Time for SW-PBIS team to meet • Time to conduct Double Check Professional Development series • Time to coach teachers			**Staff Engagement** • Enhanced efficacy to engage, teach, and respond effectively to student behavior • Reduced time spent on discipline and increased time focused on academics	**Staff Engagement** • Improvements in perceptions of school climate and organization • Reduced burnout • Increased job satisfaction • Increased staff retention
Trained Coaches • Trained in PBIS and data-based decision-making • Trained in cultural responsivity • Trained in Classroom Check-up				

Note. SW-PBIS = Schoolwide Positive Behavioral Interventions and Supports